

50  
Copies

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

*Technical Report 32-1078*

*The Enhancement of Data by Data Compression  
Using Polynomial Fitting*

*William Kizner*

FACILITY FORM 602

N67-37399	
(ACCESSION NUMBER)	(THRU)
19	1
(PAGES)	(CODE)
CI-88762	19
(NASA CR OR TMX OR AD NUMBER)	(CATEGORY)

JET PROPULSION LABORATORY  
CALIFORNIA INSTITUTE OF TECHNOLOGY  
PASADENA, CALIFORNIA

October 1, 1967

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

*Technical Report 32-1078*

*The Enhancement of Data by Data Compression  
Using Polynomial Fitting*

*William Kizner*

Approved by:

*Carlton B. Solloway for*

W. G. Melbourne, Manager  
Systems Analysis Research Section

JET PROPULSION LABORATORY  
CALIFORNIA INSTITUTE OF TECHNOLOGY  
PASADENA, CALIFORNIA

October 1, 1967

**TECHNICAL REPORT 32-1078**

Copyright © 1967  
Jet Propulsion Laboratory  
California Institute of Technology

Prepared Under Contract No. NAS 7-100  
National Aeronautics & Space Administration

## Contents

<b>I. Introduction</b> . . . . .	1
<b>II. Approximation of Functions by Polynomials</b> . . . . .	1
<b>III. The Convergence of Chebyshev Series</b> . . . . .	3
<b>IV. The Use of Convergence Properties in Estimation</b> . . . . .	7
<b>V. An Example</b> . . . . .	10
<b>VI. General Conclusion</b> . . . . .	11
<b>Appendix. The Evaluation of Chebyshev Series</b> . . . . .	12

## Tables

1. Chebyshev polynomial expansions for $e^x$ and $e^{-x}$ on interval $[0, 1]$ giving coefficients $a_n$ of $\sum a_n T_n(t)$ , $t = 2x - 1$ . . . . .	3
2. Chebyshev polynomial expansion for $\ln(1 + x)$ , $0 \leq x \leq 1$ , giving coefficients $a_n$ of $\sum a_n T_n(\tau)$ , $\tau = 2x - 1$ . . . . .	6

## Figure

1. Expansion of $\ln(1 + x) = \sum a_n T_n(\tau)$ , $0 \leq x < 1$ , $\tau = 2x - 1$ . . . . .	6
--	---

## **Abstract**

Optimal data compression by polynomial fitting is discussed from a rigorous theoretical point of view. By means of information gained from smoothness properties of the functions based on the constructive theory of functions, it is shown how estimates can be improved over the usual minimum-variance unbiased estimates. A conclusion is that measurements should be made as often as possible.

The theory developed here extends beyond the applications given to polynomial fitting. Proof is given for various theorems concerning biased estimators. One theorem, given on the result of using an approximation to the covariance matrix of the errors of the observations, is believed to be new. A similar technique might be used in estimating any set of parameters whose magnitude decreased to zero in a reasonable way.

# The Enhancement of Data by Data Compression Using Polynomial Fitting

## I. Introduction

One method of data compression used in orbit determination consists of fitting many observations of a given data type in a fixed time interval by a polynomial involving fewer parameters. Then instead of dealing with the many observations, the parameters of the polynomial are used.

It is commonly assumed that such data compression must necessarily degrade the orbit determination. Here it is shown that the usual methods of fitting polynomials, which have a bad theoretical foundation, may lead to decreased accuracies, but optimal methods should significantly increase the accuracies obtained by the usual unbiased minimum-variance formulas.

The first problem with which we will concern ourselves is that of polynomial fitting for the general case when the function to be fitted is not an exact polynomial of finite degree. Using the results of approximation theory, we show how one can take advantage of the smoothness of the function to be approximated in an optimal way by obtaining, in a sense, *a priori* values for the polynomial coefficients.

The result is an estimate of the orbital parameters that may be slightly biased, but has better accuracy than that obtained from the minimum-variance unbiased estimate without data compression. Also one need not choose a unique degree for the approximating polynomial. With the method used in this Report, the same result obtains if the degree is chosen large enough. This means that one does not have to apply the usual F test, which usually cannot be theoretically justified.

Various theorems are proved about biased estimators. One theorem, which the author believes to be new, concerns bounds for errors introduced by using a wrong noise method.

## II. Approximation of Functions by Polynomials

Engineers know that any "smooth" function  $f(x)$ ,  $a \leq x \leq b$ , can be represented by a polynomial with arbitrary accuracy in a given interval, so they often use the following simplistic method for obtaining the desired approximation. They choose an integer  $n$ , calculate a step size  $\Delta x = (b - a)/n$ , and evaluate  $f(x)$  at  $x = a, a + \Delta x, a + 2\Delta x, \dots, a + n\Delta x = b$ . A polynomial of degree  $n$ ,

$p_n(x)$ , is fitted to these points. If  $n$  is chosen high enough, then the actual result is supposed to be a polynomial  $p_m(x)$  of degree  $m$  with  $m < n$ . This is not the case, however. Such a procedure may "blow up" [ $p_n(x)$  may not converge to  $f(x)$  in any sense for increasing  $n$ ] even for functions  $f(x)$  that are infinitely differentiable. If the reader cares to verify this he can try equally spaced interpolation on  $f(x) = 1/(1+x^2)$  for the interval  $[-5, 5]$ . No matter how accurately he will do the interpolation, the error in the representation  $\max_x |f(x) - p_n(x)|$   $[-5 \leq x \leq 5]$  will increase beyond all bounds as  $n$  becomes large.

An introductory account of the constructive theory of functions, which deals with problems of approximation by polynomials, is given by Todd (Ref. 1). A more complete account, which covers the material we will use, is given by Lorentz (Ref. 2).

Some classical results that we will use are:

1. Let  $H_n$  be the set of all polynomials whose degree does not exceed  $n$ , and  $p_n(x)$  a member of  $H_n$ . For continuous  $f(x)$ ,  $a \leq x \leq b$ , define

$$E_n = \inf_{p_n \in H_n} \max_{a \leq x \leq b} |f(x) - p_n(x)|$$

Then there exists a unique member of  $H_n$ ,  $\bar{p}_n(x)$  such that  $E_n = \max_x |f(x) - \bar{p}_n(x)|$ ;  $\bar{p}_n(x)$  is called the polynomial of best approximation (in the sense of the uniform norm).

2.  $\lim_{n \rightarrow \infty} E_n = 0$  for all continuous functions.

3. The manner in which  $E_n$  approaches zero depends on the differentiability of  $f(x)$ . The set of real functions having  $k$  continuous derivatives on  $[a, b]$  is denoted by  $C^k$ . The greater the value of  $k$ , the faster  $E_n$  will approach zero with increasing  $n$ . In fact, one can deduce the value of  $k$  from the behavior of  $E_n$  as a function of  $n$ .

4. It does not make sense (in a practical sort of way) to try to accurately approximate a function with a polynomial unless it is many times differentiable, or better still, infinitely differentiable.

5. The polynomial of best approximation  $\bar{p}_n(x)$  cannot be obtained in general by linear methods. There is no way of finding  $\bar{p}_n(x)$  by evaluating  $f(x)$  at a predetermined number of values of  $x$  and calculating coefficients of the polynomial by linear relations. The algorithms for determining  $\bar{p}_n(x)$  are nonlinear and are usually time-consuming.

6. Usually acceptable polynomial approximations can be obtained by truncating a Chebyshev series expansion.

Let  $f(x)$  be defined for  $-1 \leq x \leq 1$  and satisfy a Dini-Lipschitz condition, or

$$\lim_{\delta \rightarrow 0} [\omega(\delta) \ln \delta] = 0 \quad (1)$$

where  $\omega(\delta)$  is the modulus of continuity. Let  $T_n(x)$  be the Chebyshev polynomial of the first kind of degree  $n$ . Then if  $f(x)$  satisfies the Dini-Lipschitz condition,  $f(x)$  can be represented by a uniformly convergent Chebyshev polynomial series

$$f(x) = \sum_{n=0}^{\infty} a_n T_n(x), \quad -1 \leq x \leq 1 \quad (2)$$

where

$$\left. \begin{aligned} a_0 &= \frac{1}{\pi} \int_{-1}^1 \frac{f(x)}{\sqrt{1-x^2}} dx \\ a_n &= \frac{2}{\pi} \int_{-1}^1 \frac{f(x) T_n(x)}{\sqrt{1-x^2}} dx, \quad n = 1, 2, \dots \end{aligned} \right\} \quad (3)$$

The Chebyshev polynomials are defined by

$$T_n(x) \equiv \cos(n \arccos x) \quad (4)$$

The condition (1) is satisfied if  $f(x)$  satisfies a Lipschitz condition, or if  $f'(x)$  is bounded on  $[-1, 1]$ .

For well-behaved functions the truncated expansion of Eq. (2) is a good approximation and normally has a maximum absolute error not much greater than for  $\bar{p}_n(x)$ . The drawback of this expansion is that it is usually not possible to express the  $a_n$  in terms of known functions. However, approximate numerical methods for obtaining the  $a_n$  are available.

7. The coefficients  $a_n$  for  $n = 0, 1, 2, \dots, N$  are given approximately by

$$\left. \begin{aligned} a_0 &\cong \frac{1}{m} \sum_{j=1}^m f(x_j^m) \\ a_n &\cong \frac{2}{m} \sum_{j=1}^m f(x_j^m) T_n(x_j^m), \quad n = 1, 2, \dots, N \end{aligned} \right\} \quad (5)$$

where  $m$  is an integer  $> N$  and  $x_j^m$ ,  $j = 1, 2, \dots, m$ , are the  $m$  zeros of  $T_m(x)$ . The roots are given by

$$x_j^m = \cos \frac{\pi(2j-1)}{2m}, \quad k = 1, 2, \dots, m \quad (6)$$

When  $m$  is taken to be  $N+1$ , then Eq. (5) gives us the coefficients of the interpolated  $N$ th-degree polynomial

determined by the  $N + 1$  roots of  $T_{N+1}(x)$ . It can be shown that for fixed  $N$  the approximate values of Eq. (5) will converge to the exact values given by the integrals in Eq. (3) as  $m \rightarrow \infty$  for any continuous function  $f(x)$ .

8. To establish this result we interpret Eq. (5) as the approximation of Eq. (3), using Gaussian quadrature with a weight function  $(1 - x^2)^{-1/2}$ . Here the form of the approximation is

$$\int_{-1}^1 \frac{g(x)}{\sqrt{1-x^2}} dx \cong \frac{\pi}{m} \sum_{j=1}^m g \left[ \cos \frac{\pi(2j-1)}{2m} \right] \quad (7)$$

We next use a theorem by Stieltjes, which establishes that the general Gaussian quadrature scheme (increasing  $m$ ) for any weight function on a finite interval is convergent for any continuous function  $g(x)$ . In evaluating Chebyshev series and the special series in Eq. (5), various tricks can be used (discussed in the Appendix). The important point is that Chebyshev series can be accurately and quickly evaluated.

We thus have an efficient method for determining a polynomial that uniformly approximates a function (which we assume to be at least many times differentiable) to any desired accuracy. We note that the coefficients are linear combinations of certain values of  $f(x)$ .

### III. The Convergence of Chebyshev Series

The central idea of this Report is that in orbit-determination problems, at least, we can do much better than determine the degree of a polynomial by the usual F test. Without discussing this standard method in detail, we seek to determine whether the improvement in fit that results from adding one or more parameters, or making the degree of the polynomial higher, is statistically significant.

The drawbacks of this method are many. One is that there is usually no polynomial of finite degree that will fit the data exactly even if no noise were present. Next, one does not know what significance level to use for the test. Also, one must assume that the noise is Gaussian. In addition, the test is difficult to apply when the noise is correlated.

The most serious objection to this method is that it does not consider any *a priori* knowledge about the nature of the function to be approximated. If we know that there exists a polynomial, and its approximate degree, that will fit the data "satisfactorily" we really know quite a bit about the function to be approximated. As intimated

before we really know how its coefficients approach zero. Let us illustrate with a few examples. We give in Table 1 the Chebyshev coefficients for the expansions of  $e^x$  and  $e^{-x}$  in the interval  $[0, 1]$ . Of course the variable  $t$  in  $T_n(t)$  has to be related to  $x$  by a linear transformation so that  $t = -1 \leftrightarrow x = 0$  and  $t = 1 \leftrightarrow x = 1$ .

**Table 1. Chebyshev polynomial expansions for  $e^x$  and  $e^{-x}$  on interval  $[0, 1]$  giving coefficients  $a_n$  of  $\sum a_n T_n(t)$ ,  $t = 2x - 1$**

$n$	$a_n$ for $e^x$	$a_n$ for $e^{-x}$
0	1.753387654	0.645035270
1	0.850391654	-0.312841606
2	0.105208694	0.038704116
3	0.008722105	-0.003208683
4	0.000543437	0.000199919
5	0.000027115	-0.000009975
6	0.000001128	0.000000415
7	0.000000040	-0.000000015
8	0.000000001	0.000000000

Note how the magnitude of the coefficients decreases in size approximately according to  $Kp^n$ , where  $K$  and  $p$  are constants and  $p \cong 0.1$ . In fact, for  $e^x$ ,  $0 \leq x \leq 1$ , we have for all  $n$

$$|a_n| \leq 75.59 (0.0514)^n$$

For  $e^{-x}$ ,  $0 \leq x \leq 1$ , we have for all  $n$

$$|a_n| \leq 31.68 (0.0502)^n$$

The fact that we are able to do this, and the fact that these relations hold for all  $a_n$  (which we do not show here) is no accident.

We next define analyticity on the line for functions. A function  $f(x)$  defined on  $[a, b]$  is said to be analytic on the interval if for an  $x_0 \in [a, b]$  there is a power series

$$\sum_{i=0}^{\infty} C_i(x_0) (x - x_0)^i$$

convergent for  $|x - x_0| = R$ , which represents the function at all points belonging simultaneously to  $[a, b]$  and  $(x_0 - R, x_0 + R)$ . We denote by  $A[a, b]$  the class of functions analytic on segment  $[a, b]$ . If  $R = \infty$ , the function is said to be an entire function. Then we have a result obtained by Bernstein (see Bernstein, Ref. 3, p. 110, or Lorentz, Ref. 2, p. 77, for the proof).



### Theorem 1

Let  $f(x)$  be continuous on  $[a, b]$  and  $E_n = E_n(f)$  be the best approximation to  $f(x)$  by polynomials belonging to  $H_n$ .

In this case  $f(x) \in A[a, b]$  if and only if

$$E_n < Kp^n \quad (8)$$

where  $K$  and  $p < 1$  are constants.

Moreover,  $f(x)$  is an entire function if and only if

$$\lim_{n \rightarrow \infty} (E_n)^{1/n} = 0 \quad (9)$$

We restate the theorem more precisely in the terminology of the theory of functions of a complex variable.

### Theorem 2

A. If the best approximation of the function

$$f(x) \in C[a, b] \text{ [} f(x) \text{ is continuous]}$$

is such that

$$\overline{\lim}_{n \rightarrow \infty} [E_n(f)]^{1/n} = \frac{1}{R}, \quad R > 1 \quad (10)$$

then  $f(x)$  is analytic inside an ellipse, the foci of which are the points  $a$  and  $b$ , the semiaxes having the sum  $[(b-a)/2]R$ , with a singular point lying on its boundary.

B. Conversely, if  $f(x)$  is analytic inside an ellipse whose foci are  $a$  and  $b$ , and a singular point lies on the ellipse boundary, then the best approximation by polynomials on  $[a, b]$  satisfies Eq. (10),  $[(b-a)/2]R$  being the sum of both semiaxes of the ellipse.

Thus  $p \cong 1/R$ , which in turn is determined by the location of the singularities of the function. For entire functions we can find values of  $K$  and  $p$  that satisfy the inequality (8) with  $p$  arbitrarily small.

These results carry over in the convergence of coefficients of Chebyshev series. To establish this we need some preliminary results. For simplicity assume that  $[a, b]$  has been linearly transformed into  $[-1, 1]$ . We introduce a standard mapping in the complex plane

$$z = \frac{1}{2} \left( w + \frac{1}{w} \right) \quad (11)$$

where  $z = x + iy$  and  $w = Re^{i\phi}$ . The circle in the  $w$ -plane with  $R$  constant  $> 1$  maps into the ellipse  $E_R$ ,  $(x^2/c^2) + (y^2/d^2) = 1$ , with the semiaxes  $c = (1/2)[R + (1/R)]$  and  $d = (1/2)[R - (1/R)]$ , and the foci  $\pm 1$ . For  $R = 1$ ,  $E_R$  is the interval  $[-1, 1]$  covered twice. For each point  $z$  not on  $[-1, 1]$  there is exactly one ellipse  $E_R$ ,  $R > 1$ , passing through  $z$ , which can be found from the equation

$$R = |z \pm \sqrt{z^2 - 1}| \quad (12)$$

where the sign for the square root is chosen to make the absolute value of the expression  $> 1$ . We next have a theorem derived by Bernstein (Ref. 3, p. 112) on the growth of polynomials in the complex plane.

### Theorem 3

If a polynomial  $P_n$  with complex coefficients satisfies  $|P_n(x)| \leq M$  for  $-1 \leq x \leq 1$ , then

$$|P_n(z)| \leq MR^n, \quad z \in E_R, \quad R > 1 \quad (13)$$

**Proof.** Let

$$P_n \left[ \frac{1}{2} \left( w + \frac{1}{w} \right) \right] = \frac{Q_{2n}(w)}{v^n}$$

where  $Q_{2n}(v)$  is a  $2n$ th-degree polynomial in  $v$ . For  $|v| = 1$ , using the property of the mapping (11),

$$\left| \frac{Q_{2n}(w)}{v^n} \right| = |P_n(x)| \leq M$$

Thus

$$Q_{2n}(w) \leq M \quad (14)$$

From the maximum principle the inequality (14) holds for  $|w| \leq 1$ . Let  $|w| = 1/R$ , so that  $z = (1/2)[w + (1/w)]$  lies on the ellipse  $E_R$ . Then

$$|P_n(z)| = \left| \frac{Q_{2n}(w)}{w^n} \right| \leq MR^n$$

which was to be proved.

We now can prove the following theorem about the convergence of Chebyshev series in a manner that closely follows Theorem 2.

#### Theorem 4

If the Chebyshev series expansion of  $f(x) \in C[-1, 1]$  is such that

$$\overline{\lim}_{n \rightarrow \infty} \left( |a_n| \right)^{1/n} = \frac{1}{R_0}, \quad R_0 > 1 \quad (15)$$

then  $f(x)$  is analytic inside the ellipse  $E_{R_0}$ , with a singular point lying on  $E_{R_0}$ . Conversely, if  $f(x)$  is analytic inside  $E_{R_0}$  for  $R_0 > 1$ , then the coefficients satisfy Eq. (15).

**Proof.** Let  $D_R$  be the interior of the ellipse  $E_R$ . We show first that if  $f$  is analytic in  $D_R$ ,  $1 < R < \infty$ , then

$$\overline{\lim}_{n \rightarrow \infty} \left( |a_n| \right)^{1/n} \leq \frac{1}{R} \quad (16)$$

We expand  $f(x)$  on  $[-1, 1]$  in Chebyshev polynomials according to Eq. (2) and (3); the fact that  $f(x)$  has a continuous derivative assures us that the series is uniformly convergent. By the substitution  $\cos t = x$  we obtain

$$\begin{aligned} f(\cos t) &= \sum_{k=0}^{\infty} a_k \cos k t \\ a_0 &= \frac{1}{2\pi} \int_{-\pi}^{\pi} f(\cos t) dt \\ a_n &= \frac{1}{\pi} \int_{-\pi}^{\pi} f(\cos t) \cos k t dt, \quad n = 1, 2, \dots \end{aligned}$$

The substitution  $z = e^{it}$  in the integral for  $a_n$  gives the line integral along the circle  $|z| = 1$ ,

$$a_n = \frac{1}{\pi i} \int_C f\left(\frac{z+z^{-1}}{2}\right) \frac{z^n + z^{-n}}{2} \frac{dz}{z}, \quad n = 1, 2, \dots \quad (17)$$

Choose  $R_1$  with  $1 < R_1 < R$ . Consider the function

$$g(z) = f\left(\frac{z+z^{-1}}{2}\right)$$

in the closed ring  $G$  bounded by the circles  $C_1: |z| = R_1^{-1}$ , and  $C_2: |z| = R_1$ . From the discussion following Eq. (11) we see that if  $z$  is on either of the circles  $|z| = \sigma$  or  $|z| = \sigma^{-1}$ , then  $w = (1/2)(z + z^{-1})$  is on the ellipse  $E_\sigma$ . Since  $f(w)$  is analytic in  $D_R$ , it follows that  $g(z)$  is analytic in the ring  $G$ . We now change the path of integration to estimate the  $a_n$ . We split the integral (17) into two parts; the integral containing  $z^{-n}$  is taken over a circle with a

large radius, and the integral with  $z^n$  is taken over a circle with a small radius. Thus

$$a_n = \frac{1}{2\pi i} \int_{C_1} g(z) z^{n-1} dz + \frac{1}{2\pi i} \int_{C_2} g(z) z^{-(n+1)} dz$$

Let  $M$  be the maximum value of  $|f|$  on  $E_{R_1}$ . Then the absolute value of the first integral is bounded by

$$\frac{1}{2\pi} M \left( \frac{1}{R_1} \right)^{n-1} 2\pi R_1^{-1} = M R_1^{-n}$$

In the same way the second integral is bounded by  $M R_1^{-n}$ . Thus

$$|a_n| \leq \frac{2M}{R_1^n}, \quad n = 1, 2, \dots \quad (18)$$

It follows that

$$(|a_n|)^{1/n} \leq (2M)^{1/n} 1/R_1, \text{ and } \overline{\lim}_{n \rightarrow \infty} (|a_n|)^{1/n} \leq 1/R_1.$$

Since  $R_1$  can be taken arbitrarily close to  $R$ , we obtain

$$\overline{\lim}_{n \rightarrow \infty} \left( |a_n| \right)^{1/n} \leq \frac{1}{R} \quad (19)$$

Conversely, let the inequality (16) hold for some  $R$ ,  $1 < R < \infty$ . We will show that  $f$  has an analytic extension onto each elliptic disc  $D_{R_1}$ ,  $1 < R_1 < R$ , and therefore onto  $D_R$ . This taken with Eq. (19) will show that the largest elliptic disc in which  $f$  is analytic is  $D_{R_0}$ , which will prove the theorem.

Let  $R_1 < R_2 < R$ . Then  $(|a_n|)^{1/n} \leq 1/R_2$  for  $n \geq N$  with  $N$  sufficiently large. It follows that on  $[-1, 1]$

$$|a_n T_n(x)| \leq |a_n| \leq \left( \frac{1}{R_2} \right)^n, \quad n > N \quad (20)$$

where we use the fact that  $|T_n(x)| \leq 1$  for  $x \in [-1, 1]$ . Applying Theorem 3 to expression (20), using  $z$  in  $D_{R_1}$ ,

$$|a_n T_n(z)| \leq \left( \frac{R_1}{R_2} \right)^n, \quad z \in D_{R_1}, \quad n > N \quad (21)$$

Hence the Chebyshev series expansion of  $f(x)$ , with  $x$  replaced by  $z$ ,  $f(z) = \sum_{n=0}^{\infty} a_n T_n(z)$ , converges uniformly on each compact subset of  $D_R$ . Its sum is analytic on  $D_R$  and provides the desired analytic extension of  $f$ . This completes the proof.

The importance of Theorem 4 is that it shows that for functions analytic on the line there exist bounds for the Chebyshev coefficients of the form

$$|a_n| \leq Kp^n \quad (22)$$

where  $K$  and  $p < 1$  are constants, and  $p$  is arbitrarily close to  $1/R_0$ , given by Theorem 4.

Consider for example the expansion of  $f(x) = 4/(5 + 4x)$ , where  $-1 \leq x \leq 1$ . Here there is a pole at  $-5/4$ . From Eq. (12),  $R = |-5/4 - \sqrt{9/16}| = 2$ . Hence we expect

$$|a_n| \leq K \left(\frac{1}{2}\right)^n$$

for large  $n$ .

In fact, Elliott (Ref. 4) shows that for all  $n$

$$a_n = \frac{8(-1)^n}{3} \left(\frac{1}{2}\right)^n$$

Another example that shows the regularity with which  $a_n \rightarrow 0$  is given by the expansion of  $\ln(1+x)$ ,  $0 \leq x \leq 1$ , presented in Table 2.

Figure 1 shows the behavior of  $\log a_n$  vs  $n$ . In addition,  $\log(Kp^n)$  vs  $n$  is graphed, which is a straight line. Here

$$R = |-3 - \sqrt{3^2 - 1}| = 5.8284 \dots$$

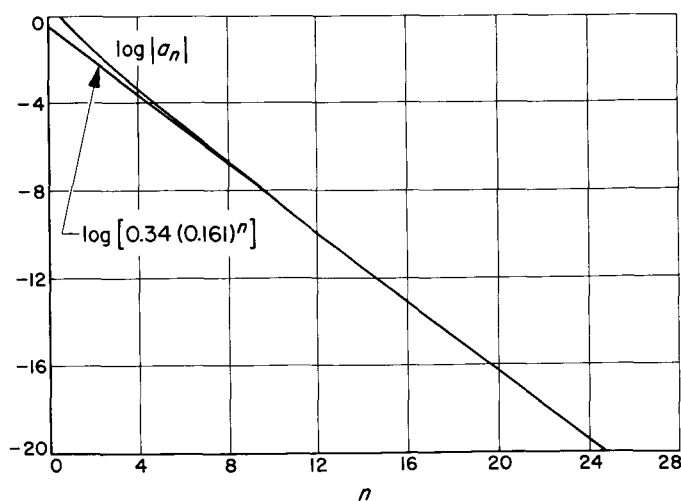


Fig. 1. Expansion of  $\ln(1+x) = \sum a_n T_n(\tau)$ ,  $0 \leq x < 1$ ,  $\tau = 2x - 1$

Table 2. Chebyshev polynomial expansion for  $\ln(1+x)$ ,  $0 \leq x \leq 1$ , giving coefficients  $a_n$  of  $\sum a_n T_n(\tau)$ ,  $\tau = 2x - 1$

$n$	$a_n$
0	$0.376 \times 10^0$
1	$0.343 \times 10^0$
2	$-0.294 \times 10^{-1}$
3	$0.336 \times 10^{-2}$
4	$-0.433 \times 10^{-3}$
5	$0.594 \times 10^{-4}$
6	$-0.850 \times 10^{-5}$
7	$0.125 \times 10^{-5}$
8	$-0.187 \times 10^{-6}$
9	$0.286 \times 10^{-7}$
10	$-0.442 \times 10^{-8}$
11	$0.689 \times 10^{-9}$
12	$-0.108 \times 10^{-9}$
13	$0.171 \times 10^{-10}$
14	$-0.273 \times 10^{-11}$
15	$0.438 \times 10^{-12}$
16	$-0.704 \times 10^{-13}$
17	$0.113 \times 10^{-13}$
18	$-0.184 \times 10^{-14}$
19	$0.299 \times 10^{-15}$
20	$-0.488 \times 10^{-16}$
21	$0.798 \times 10^{-17}$
22	$-0.131 \times 10^{-17}$

since the singularity (in the scale we use) occurs at  $Z = -3$ . The value for  $p$  used for the straight line in Fig. 1 is 0.161 and is not exactly equal to  $1/R$ , which is  $0.1715 \dots$ . Since only a finite number of terms was used, one can expect some discrepancy.

This raises the question of whether or not we can be sure that the function we are approximating is analytic on the line, since in practice we do not have an analytic expression for  $f(x)$ . In many cases we can find a differential equation for the quantity to be approximated and infer the analyticity from the nature of the equation [see for instance Lefschetz (Ref. 5)]. If the function does seem poorly behaved in a small region, it is best to eliminate this region from the approximation by using a smaller interval.

The author has generated tens of thousands of Chebyshev series (Ref. 6) and can report from the many cases he has examined that a very simple convergence pattern always holds for rapidly converging series. In fact, for functions analytic on the line,

$$|a_n| = Kp^n \quad (23)$$

often seems to hold for large  $n$ , when the function is not an entire function. In other words, an expression of the form  $Kp^n$  is not merely a good bound, but often is a good approximation for  $|a_n|$  when  $n$  is large. This may be partly justified from the results of Elliott (Ref. 4), where sharp estimates of  $a_n$  are obtained.

We shall give a simplified description of some of Elliott's results and refer the reader to Ref. 4 for exact statements. Elliott discusses the contributions to the Chebyshev coefficients of poles and a branch point on the real axis by evaluating a contour integral. The result is that the contribution to  $a_n$  is expressed as residues of functions involving  $f(z)$  evaluated at the singularities of  $f(z)$ . For large  $n$  we need only consider the effects of those singularities located on  $E_R$ , where  $R > 1$  is chosen so that there is at least one singularity of  $f(z)$  on  $E_R$  and  $f(z)$  is analytic in the interior of  $E_R$ . If the singular points on  $E_R$  consist of a single pole of order  $k$ , then Elliott shows that for large  $n$

$$|a_n| \cong K \left( \frac{1}{R} \right)^n \frac{(n+k-1)!}{n!} \quad (24)$$

where  $K$  is a constant. If  $k = 1$  then Eq. (24) is of the form (23). The Schwarz principle of reflection applied to  $f(z)$  [ $f(z)$  is real for real  $z$ ] shows us that its poles must occur at conjugate complex values of  $z$ . If we consider pairs of poles (or any other singularity) on  $E_R$ , we find that their individual effects are conjugate complex values of each other. Their effects may be characterized by a slowly varying term similar to Eq. (24) multiplied by a rapidly varying term of maximum amplitude one.

Similar results hold for the case of a branch point on the real axis. Thus when there is only one singularity on  $E_R$ , Eq. (24) is a good approximation to  $|a_n|$ , which behaves much like Eq. (23) for a restricted range of  $n$ .

Our main hypothesis is that we can use the inequality (22) by estimating  $K$  and  $p$  from a finite number of terms. The effect of errors caused by taking a finite number of terms will be discussed in connection with a general algorithm for implementing this curve-fitting method.

#### IV. The Use of Convergence Properties in Estimation

Let us first recall the usual minimum-variance linear unbiased estimator (see Scheffé, Ref. 7). Let the  $n$ -dimensional column vector  $y$  represent the observations, and

$$y = Uq + e \quad (25)$$

where  $U$  is a known  $n \times m$  matrix,  $q$  is an  $m$ -dimensional column vector representing the unknown elements we wish to estimate, and  $e$  is an  $n$ -dimensional column noise vector. We assume that the expectation of  $e$  is zero,

$$E(e) = 0 \quad (26)$$

and assume that the covariance matrix of the noise

$$E(ee^T) = \Lambda_e \quad (27)$$

is known and is positive definite. Here the superscript  $T$  indicates transpose. If the columns of  $U$  are linearly independent, by the Gauss-Markoff theorem we have the unbiased minimum-variance estimate of  $q$ ,

$$\hat{q} = (U^T \Lambda_e^{-1} U)^{-1} U^T \Lambda_e^{-1} y \quad (28)$$

The covariance matrix of  $\hat{q}$  is given by

$$E[(\hat{q} - q)(\hat{q} - q)^T] = (U^T \Lambda_e^{-1} U)^{-1} \quad (29)$$

We now show how to obtain "better" estimates by making use of the convergence properties of Chebyshev series. Since the new estimates may be biased, we define a "best estimate" to be a linear minimum second-moment estimate of  $q$ ,  $\tilde{q}$ , where the second moments are taken with respect to the true value of  $q$ . Since ordinarily it is the size of the errors in one experiment that matters, and not whether under some hypothetical situation the expectation of the errors over the hypothetical ensemble is zero, we dispense with the condition of no bias. Thus we seek a linear estimate  $\tilde{q}$  such that its second-moment matrix

$$E[(\tilde{q} - q)(\tilde{q} - q)^T] = \Lambda_{\tilde{q}} \quad (30)$$

is a minimum in some sense.

In connection with the minimization of matrices we review the following definitions. A square matrix  $A$  is called positive definite if  $x^T A x > 0$  for all  $x \neq 0$ . If  $x^T A x \geq 0$  for all  $x$  and there exists a nonzero vector  $x$  for which the equality holds,  $A$  is called positive semidefinite. By definition, a positive definite matrix is less than or equal to a second positive definite matrix if the second matrix minus the first matrix is positive semidefinite.

We now state the mathematical problem we shall solve. As before we are given a model for a set of observations  $y$ , as in Eq. (25), with  $U$  and  $y$  known and with certain known

statistics about the error vector. However, we now no longer assume that  $E(e) = 0$ , but

$$E(e) = \epsilon \quad (31)$$

where  $\epsilon$  is unknown.

Again,

$$E(ee^T) = \Lambda_e \quad (27)$$

is assumed to be positive definite. The problem is to choose a linear estimator  $\tilde{q}$  from the class of linear estimators such that  $\Lambda_{\tilde{q}} \equiv E(\tilde{q} - q)(\tilde{q} - q)^T \leq \Lambda_{\bar{q}}$ , where  $\bar{q}$  is any other linear estimator. Let

$$\bar{q} = Qy \quad (32)$$

where  $Q$  is a given matrix. Hence

$$\begin{aligned} \Lambda_{\bar{q}} &= E\{[Q(Uq + e) - q][Q(Uq + e) - q]^T\} \\ &= (QU - I)q q^T (QU - I)^T + (QU - I)\epsilon \epsilon^T Q^T \\ &\quad + Q\epsilon(U^T Q^T - I) + Q\Lambda_e Q^T \end{aligned} \quad (33)$$

where  $I$  is the unit matrix. This results in  $\Lambda_{\bar{q}}$  being dependent on the unknown  $q$  and  $\epsilon$ . If we restrict our estimators by the condition

$$QU = I \quad (34)$$

which previously resulted from the condition of no bias, then

$$\Lambda_{\bar{q}} = Q\Lambda_e Q^T \quad (35)$$

no longer depends on the unknown  $q$  or  $\epsilon$ . This means that the solution to the problem of finding an estimator with a minimum  $\Lambda_{\tilde{q}}$  reduces to the form of the previous case, with

$$\tilde{q} = (U^T \Lambda_e^{-1} U)^{-1} U^T \Lambda_e^{-1} y \quad (36)$$

where the tilde is used to remind us that  $E(q - q) \neq 0$ . In fact, using Eq. (34) and (36),

$$E(\tilde{q} - q) = (U^T \Lambda_e^{-1} U)^{-1} U^T \Lambda_e^{-1} \epsilon \quad (37)$$

We next develop a special formula for the use of *a priori* information with bias. The most general formula is left as an exercise for the reader. We reinterpret the observation vector to include the *a priori* estimate of  $q$ ,  $q_0$  for the first part of  $y$ . Thus  $y'$  is partitioned as

$$y' = \begin{pmatrix} q_0 \\ y \end{pmatrix} \quad (38)$$

By definition

$$U' = \begin{pmatrix} I \\ U \end{pmatrix} \quad (39)$$

is the augmented  $U$  matrix. We consider the special case where the augmented  $\Lambda_e$  matrix is of the form

$$\Lambda' = \begin{pmatrix} \Lambda_0 & 0 \\ 0 & \Lambda_e \end{pmatrix} \quad (40)$$

When Eq. (36) is evaluated with the new quantities, certain simplifications result:

$$\tilde{q} = (\Lambda_0^{-1} + U^T \Lambda_e^{-1} U)^{-1} (\Lambda_0^{-1} U^T \Lambda_e^{-1}) y' \quad (41)$$

and

$$\Lambda_{\tilde{q}} = (\Lambda_0^{-1} + U^T \Lambda_e^{-1} U)^{-1} \quad (42)$$

We now have the basic mathematical machinery to make use of convergence properties of series. We wish to estimate  $A$ , the vector of coefficients for the Chebyshev polynomial series. Without *a priori* information

$$y = BA + e \quad (43)$$

where the  $B$  matrix is known in terms of the Chebyshev polynomials. For programming purposes it is convenient to have indices start at one, instead of zero. Hence,

$$B_{ij} = T_{j-1}(\tau_i), \quad \begin{matrix} i = 1, 2, \dots, n \\ j = 1, 2, \dots, L \end{matrix} \quad (44)$$

where  $n$  is the number of observations of the particular data type, or of the curve we are fitting,  $L$  is the dimension of  $A$ , and  $\tau_i$  is the transformed time (to  $[-1, 1]$ ) of the  $i$ th observation. For the moment let us choose  $L$  to be so large that the error in the representation of  $E(y)$  is negligible. Mathematically there is no limit to how large  $L$  might be chosen, although if  $L > n$  we have to introduce some complicated algebraic notions (such as generalized inverses). If the interval on which the observations are given is denoted by  $[t_0, t_f]$ , then

$$\tau_i = \frac{2}{t_f - t_0} \left( t_i - \frac{t_0 + t_f}{2} \right), \quad i = 1, 2, \dots, n \quad (45)$$

The evaluation of a Chebyshev polynomial for Eq. (44) may be regarded as a special case of the evaluation of a Chebyshev series, which is discussed in the Appendix.

Now we assume that we know that the series is bounded by an expression such as (22). Then there exists an *a priori* estimate of  $A$ , which we take to be zero, and a finite second-moment matrix of this *a priori* estimate, which we call  $\Lambda_{a.p.a.}$ .

It is easily shown that the best second-moment matrix for the *a priori* estimate depends on the unknown  $A$ . The matrix is given by  $AA^T$ , which obviously is positive semi-definite. Since the value of  $A$  is unknown, we might approximate  $AA^T$  in some manner by using the available observations and any other *a priori* information. Thus we will have to consider the effect of using an approximate value for  $\Lambda_{a.p.a.}$ . This will be done in a more general context when we consider the effect of using a wrong second-moment matrix of the noise. For simplicity we assume that the approximate  $\Lambda_{a.p.a.}$  that is used is positive definite. If we assume that the true value of each component of  $A$  is uncorrelated with the error in any component of observation other than  $A$ , or

$$E(0 - a_i)(e_j) = 0, \quad \begin{matrix} i = 1, 2, \dots, L \\ j = L + 1, \dots, n + L \end{matrix} \quad (46)$$

then we can partition  $\Lambda'$  as in Eq. (40) and obtain

$$\tilde{A} = (\Lambda_{a.p.a.}^{-1} + B^T \Lambda_e^{-1} B) B^T \Lambda_e^{-1} y \quad (47)$$

where use has been made of the fact that the *a priori* estimate of  $A$  is taken to be zero.

For the case of orbit determination we use  $\tilde{A}$  as a means of estimating  $q$ , the orbital parameters.

If

$$A = Cq \quad (48)$$

then

$$\tilde{q} = [C^T \Lambda_a^{-1} C]^{-1} C^T \Lambda_a^{-1} \tilde{A} \quad (49)$$

where

$$\Lambda_a^{-1} = (\Lambda_{a.p.a.}^{-1} + B^T \Lambda_e^{-1} B) \quad (50)$$

From Eq. (5), (6), and (48), we have

$$\left. \begin{aligned} C_{ij} &= \frac{1}{L+1} \sum_{k=1}^{L+1} \frac{\partial f(t_k)}{\partial q_j}, & i &= 1 \\ & & j &= 1, 2, \dots, m \\ C_{ij} &= \frac{2}{L+1} \sum_{k=1}^{L+1} \frac{\partial f(t_k)}{\partial q_j} T_{i-1}(\tau_k), & i &= 2, 3, \dots, L \\ & & j &= 1, 2, \dots, m \end{aligned} \right\} \quad (51)$$

where  $m$  is the dimension of  $q$ . Here we assume that the dimension of  $q$  is less than that of  $A$  and the indicated inverses exist. We assume that  $f(t)$  is given in the form

$$f(t) = \sum_{j=1}^m f_j(t) q_j$$

Also

$$\left. \begin{aligned} \tau_k &= \cos \frac{\pi}{2} \left( \frac{2k-1}{L+1} \right) \\ t_k &= \frac{t_0 + t_f}{2} + \frac{t_f - t_0}{2} \tau_k \end{aligned} \right\} \quad (52)$$

If we combine Eq. (49) and (47),

$$\tilde{q} = (C^T \Lambda_a^{-1} C)^{-1} C^T B^T \Lambda_e^{-1} y \quad (53)$$

where the largest matrix to be inverted (besides  $\Lambda_e$ , which requires special treatment) is

$$\Lambda_{\tilde{q}}^{-1} = (C^T \Lambda_a^{-1} C) \quad (54)$$

instead of  $\Lambda_a^{-1}$  as in Eq. (47) and (49). The disadvantage of Eq. (53) is that it deals with coordinate deviations directly and the data are not "compressed." Note that in Eq. (53)  $C^T B^T$  is simply the transpose of the matrix  $U$ , which relates the observations to the orbital elements as in Eq. (25).

We now discuss the consequences of the assumptions or approximations that we are making. One assumption, that the columns of the  $U$  matrix are linearly independent, can be dealt with very simply. This is discussed by Scheffé (Ref. 7). The result is that if the columns are in fact linearly dependent, then the minimum-variance estimates are no longer unique as far as their representation in terms of the parameters go. However, they are unique in terms of the residual quadratic form. For the orbit-determination problem this means that if we try to determine too many astronomical constants from a set of observations of a space

probe, the result will be failure to determine the constants uniquely, but nevertheless the residuals of the observations will be uniquely determined. Thus if one is primarily interested in orbit prediction for short times in the future, it should not matter whether or not the columns of the  $U$  matrix are linearly dependent. To simplify the discussion we can always eliminate some parameters to make the columns linearly independent.

Our next topic is the effect of using approximations for  $\Lambda_e$ , in particular for  $\Lambda_{a.p.a.}$ , which forms a part of  $\Lambda_e$ . A general discussion of the approximation of using a diagonal matrix for  $\Lambda_e$  is given by Magness and McGuire (Ref. 8). A question they ask is not how much better one can do with a minimum-variance estimate, since this is unattainable when the cross correlations are unknown, but whether the unknown correlations "hurt" you. Unfortunately, one has to go through certain transformations to find out the effect of the unknown correlations.

Here we take a different approach.

#### Theorem 5

Given a model for the observations  $y = Uq + e$ , where the columns of  $U$  are linearly independent and  $E(ee^T) = \Lambda_e$  is positive definite,  $E(e) = \epsilon$ , which may not be zero. Suppose we construct an estimator for  $q$  according to Eq. (36), but use an incorrect value of  $\Lambda_e$ , which we denote by  $\Gamma$ , and assume that  $\Gamma$  is positive definite. Then the matrix multiplying the observations is

$$Q = (U^T \Gamma^{-1} U)^{-1} U^T \Gamma^{-1} \quad (55)$$

Let  $\Gamma \geq \Lambda_e$ , or  $\Gamma - \Lambda_e = S$ , where  $S$  is positive semidefinite. Then the actual second-moment matrix of the error in  $\tilde{q}$  will be less than or equal to the nominal one,  $(U^T \Gamma^{-1} U)^{-1}$ .

**Proof.** We recall the result from matrix theory that  $U^T A U$  is positive definite if  $A$  is positive definite and the columns of  $U$  are linearly independent. Hence  $(U^T \Gamma^{-1} U)^{-1}$  is positive definite. Also, the columns of  $Q^T$  are linearly independent by the construction of Eq. (55). Hence the actual second-moment matrix, from Eq. (35), is

$$\Lambda_{act} = Q \Lambda_e Q^T \leq Q (\Lambda_e + S) Q^T \quad (56)$$

The inequality in Eq. (56) certainly holds if  $S$  is positive definite. If it is positive semidefinite, the inequality can be

proved from the continuity of the product as a function of  $S$ . Also

$$Q (\Lambda_e + S) Q^T = Q \Gamma Q^T = (U^T \Gamma^{-1} U)^{-1} \quad (57)$$

using Eq. (55).

**Corollary.** The same result holds if  $\Lambda_e$  is positive semidefinite and  $\Gamma = S + \Lambda_e$ , where  $\Gamma$ , the nominal-error second-moment matrix, is positive definite and  $S$  is positive semidefinite.

Equations (34) and (35) will still hold and Eq. (56) and (57) follow as before, which proves the corollary.

**Corollary.**  $\Gamma$  [used in Eq. (55) and (57)] may be chosen to be a diagonal matrix.

It is easily shown that if the elements of the diagonal matrix  $\Gamma$  are made large enough, then  $\Gamma - \Lambda_e$  is a positive semidefinite matrix. In fact, if each element of  $\Gamma$  is larger than the largest eigenvalue of  $\Lambda_e$ , the corollary holds.

Incidentally, it is easy to obtain a simple bound for the largest eigenvalue of a matrix  $\Lambda$ . For instance, it can be shown that the largest eigenvalue,  $\lambda_{max}$ , is bounded by

$$\lambda_{max} \leq \max_j \left( \sum_{k=1}^n \left| \Lambda_{jk} \right| \right)$$

where  $n$  is the dimension of  $\Lambda$ . (This is a modification of a theorem derived by L. Lévy.)

#### V. An Example

In Table 1 we gave the Chebyshev expansions for  $e^x$  and  $e^{-x}$  on the interval  $[0, 1]$ . Suppose we are trying to estimate the function  $f(x) = q e^{-x}$  on the same interval  $[0, 1]$  where  $q$  is unknown. Assume that in reality  $q = 1$ . Assume that we have some knowledge of the size of  $a_i$ , which was obtained from a study of comparison functions. For simplicity we will limit the dimension of  $A$  to 3, or assume that we adequately approximate the function with a polynomial of degree  $3 - 1 = 2$ . We also assume that we are limited to three measurements at time  $x = 0$ ,  $\frac{1}{2}$ , and 1.

If we restrict ourselves to these three time points, it can easily be shown that  $e^{-x}$  evaluated at these points determines a second-degree polynomial with slightly different

coefficients than given by Table 1. This is obtained by evaluating  $e^{-x}$  at these three time points, 0,  $\frac{1}{2}$ , and 1 (1, 0.60653, 0.36788 to five decimal places) and interpolating. The resulting series is

$$f(\tau) = 0.645235 - 0.31606 T_1(\tau) + 0.038705 T_2(\tau) - 1 \leq \tau \leq 1 \quad (58)$$

Next, assume that the covariance matrix for  $A$  that is available to us is 10 times the squares of the values for  $a_i$  given in Table 1. The factor of 10 provides us with a margin of safety for our estimating method:

$$\Lambda_{a.p.a.} = \begin{pmatrix} 4.16025 & 0 & 0 \\ 0 & 0.97969 & 0 \\ 0 & 0 & 0.01521 \end{pmatrix} \quad (59)$$

$$\Lambda_{a.p.a.}^{-1} = \begin{pmatrix} 0.24037 & 0 & 0 \\ 0 & 1.02073 & 0 \\ 0 & 0 & 65.7462 \end{pmatrix} \quad (60)$$

Also assume that

$$\Lambda_e = I, \quad E(e) = 0 \quad (61)$$

These are our basic assumptions. Next, we use the usual unbiased minimum-variance estimate given by Eq. (28) with the covariance matrix given by Eq. (29).

Here

$$U^T = (1, 0.60653, 0.36788) \quad (62)$$

Then the variance of the estimate given by Eq. (28) is, from Eq. (29),

$$(U^T U)^{-1} = 0.66524 \quad (63)$$

We now redo the problem, using our *a priori* knowledge of the size of the coefficients, and see what we gain. From Eq. (44) and a knowledge of the behavior of Cheby-

shev polynomials,

$$B = \begin{pmatrix} 1 & -1 & 1 \\ 1 & 0 & -1 \\ 1 & 1 & 1 \end{pmatrix} \quad (64)$$

$$\Lambda_a^{-1} = B^T B + \Lambda_{a.p.a.}^{-1} = \begin{pmatrix} 3.24037 & 0 & 1 \\ 0 & 3.02037 & 0 \\ 1 & 0 & 68.7462 \end{pmatrix} \quad (65)$$

From Eq. (58) and (51),

$$C^T = (0.645235, -0.31606, 0.038705) \quad (66)$$

$$\Lambda_{\tilde{q}}^{-1} = C^T \Lambda_a^{-1} C = 1.8037447 \quad (67)$$

$$\Lambda_{\tilde{q}} = 0.5544022 \quad (68)$$

Finally, from Eq. (53),

$$\tilde{q} = (0.5544022, 0.3362614, 0.2039532) y \quad (69)$$

We note that Eq. (68) gives the estimated variance of  $\tilde{q}$ . The actual variance of  $q$  is obtained by using Eq. (69) and the statistics of  $y$ . The result is a variance of 0.462030 plus a bias term of 0.166661. The total second moment about the true value is

$$0.462030 + (0.166661)^2 = 0.48979$$

which is smaller than that given by Eq. (63) (0.66524), using the usual formula. The reason that the actual second moment is smaller than the calculated one given by Eq. (63) is that we allowed ourselves plenty of margin for error in estimating the size of  $a_n$ .

## VI. General Conclusion

The method we used to improve our polynomial fitting can be used in the estimation of other parameters if we give up the condition of unbiasedness. Equation (28) is the best linear unbiased estimate. If we can determine bounds for  $E(qq^T)$ , then we can employ the same procedure and use an *a priori* estimate of  $q$  as zero, with the associated second-moment matrix.



## Appendix

### The Evaluation of Chebyshev Series

The evaluation of a series of orthogonal polynomials can be conveniently accomplished by a method described in Ref. 6. For a Chebyshev polynomial series the algorithm is as follows:

Given the series

$$p(x) = \sum_{n=0}^N a_n T_n(x)$$

Define

$$b_{n+1} = b_{n+2} = 0$$

Compute recursively

$$b_n = 2x b_{n+1} - b_{n+2} + a_n, \quad n = N, N-1, \dots, 1$$

$$b_0 = x b_1 - b_2 + a_0$$

$$p(x) = b_0$$

These formulas may also be used in curve fitting as in Eq. (5) and (6). If maximum accuracy is desired for curve fitting, another method is recommended. This alternative method is most useful when  $n$  is large and the round-off error in  $x$ , as used in the above algorithm, may be significant.

To obtain the Chebyshev coefficients in curve fitting, we make use of the definition of  $T_n(x)$  as given by Eq. (4) when used in Eq. (5) and (6), which results in

We now observe that the cosine function need only be evaluated in the first quadrant for the values

$$\cos \frac{\pi}{2} \frac{k}{m}, \quad k = 0, 1, \dots, m$$

For values of  $n(2j-1)$  that are greater than  $m$ , the quadrant  $L$  is found from

$$L = \left[ \frac{n(2j-1)}{m} \right] \left[ \text{mod } 4 \right] + 1$$

where integer arithmetic is used for the first division. Then the well-known formulas for transforming the cosine function to a function of the first quadrant are used to obtain

$$k_0 \equiv n(2j-1) \pmod{m}$$

$$T_n(x_j^m) = \cos \frac{\pi}{2} \frac{k_0}{m}, \quad L = 1$$

$$T_n(x_j^m) = -\cos \frac{\pi}{2} \frac{k}{m}, \quad L = 2$$

where  $k = m - k_0$ , and

$$T_n(x_j^m) = -\cos \frac{\pi}{2} \frac{k_0}{m}, \quad L = 3$$

$$T_n(x_j^m) = \cos \frac{\pi}{2} \frac{k}{m}, \quad L = 4$$

where  $k = m - k_0$ .

---


$$a_n \cong \frac{2}{m} \sum_{j=1}^m f(x_j^m) T_n \left[ \cos \frac{\pi}{2} \frac{(2j-1)}{m} \right], \quad n = 1, 2, \dots, N$$

$$= \frac{2}{m} \sum_{j=1}^m f(x_j^m) \cos \left[ \frac{\pi}{2} \frac{n(2j-1)}{m} \right], \quad n = 1, 2, \dots, N$$

## References

1. Todd, J., *Introduction to the Constructive Theory of Functions*, Academic Press, Inc., New York, 1963.
2. Lorentz, G. G., *Approximation of Functions*, Holt, Rinehart and Winston, Inc., New York, 1966.
3. Bernstein, S., *Leçons sur les Propriétés Extrémales et la Meilleure Approximation des Fonctions Analytiques d'une Variable Réelle*, Gauthier-Villars, Paris, 1926.
4. Elliott, E., "The Evaluation and Estimation of the Coefficients in the Chebyshev Series Expansion of a Function," *Mathematics of Computation*, Vol. 18, No. 86, p. 274, 1964.
5. Lefschetz, S., *Differential Equations: Geometric Theory*, 2nd edition, Interscience Publishers, New York, p. 43, 1957.
6. Kizner, W., *Error Curves for Lanczos' "Selected Points" Method*, Technical Report 32-703, Jet Propulsion Laboratory, Pasadena, Calif., April 30, 1966 (Reprint from *Computer Journal*, Vol. 8, No. 4, pp. 372-382, January 1966).
7. Scheffé, H., *The Analysis of Variance*, Ch. 1, John Wiley & Sons, Inc., New York, 1959.
8. Magness, T. A., and McGuire, J. B., "Comparison of Least Squares and Minimum Variance Estimates of Regression Parameters," *Annals of Mathematical Statistics*, Vol. 33, pp. 462-470, 1962.